|  |  |  |
| --- | --- | --- |
| **A** | **Course Title  & Number** | **Advanced data mining: theory and applications - MTH 594** |
| **B** | **Pre/Co-requisite(s)** | Approval of the MSMTH graduate program coordinator |
| **C** | **Number of credits** | 3-0-3 |
| **D** | **Faculty Name** | **Dr. Dmitry Efimov** |
| **E** | **Term/ Year** | Spring 2016 |
| **F** | **Sections** | |  |  |  |  |  | | --- | --- | --- | --- | --- | | **CRN** | **Course** | **Days** | **Time** | **Location** | | 21576.01 | MTH 594 | Saturday | 3.00 - 5:45 pm | NAB 1 008 | |  | | | | | |
| **G** | **Instructor Information** | |  |  |  |  | | --- | --- | --- | --- | | **Instructor** | **Office** | **Telephone** | **Email** | | Dmitry Efimov | NAB127 | 515-4458 | defimov@aus.edu |   Office Hours:   |  |  |  |  |  | | --- | --- | --- | --- | --- | | **Sunday** | **Monday** | **Tuesday** | **Wednesday** | **Thursday** | | 11am-1pm |  | 11am-1pm |  |  | |
| **H** | **Course Description from Catalog** | Covers foundation and theory of supervised and unsupervised machine learning methods including linear and logistic regressions, regularization, artificial neural networks, support vector machines, decision trees, ensembling, k-means and hierarchical clustering, principal component analysis, recommender systems, Naive Bayes. The course also covers some related recent research topics. |
| **I** | **Course Learning Outcomes** | Upon completion of the course, students will be able to:   1. Demonstrate understanding of the theoretical aspects of linear and logistic regressions, regularization, multilayer neural networks, support vector machine, kernel functions, decision trees, clustering. 2. Apply linear and logistic regression with regularization for different datasets. 3. Construct multilayer neural networks, support vector machine with different kernel functions and decision trees with entropy and Gini splitting criteria. 4. Implement unsupervised learning algorithms such as k-means, hierarchical clustering and principal component analysis. 5. Demonstrate understanding of simple algorithms for recommender systems and their applications. 6. Solve machine learning problems using basic probabilistic approach Naive Bayes. 7. Conduct independent work on projects related to data analysis, prepare written reports, and present and defend the results in public presentations. |
| **J** | **Textbook and other Instructional Material and Resources** | There is no textbook for this course. Lecture notes, handouts, and selected readings from material available in the library will be used. The following books are good resources and are available in the University library:   1. T.Hastie, R.Tibshirani and J.Friedman, The elements of statistical learning: Data Mining, Inference, and Prediction, 2nd edition, Springer, 2009. 2. S.Shalev-Shwartz, S.Ben-David, Understanding Machine Learning: from theory to algorithms, Cambridge University Press, 2014. |
| **K** | **Teaching and Learning Methodologies** | Lectures, class discussions including case studies. |
| **L** | **Grading Distribution, and Due Dates** | **Grading Distribution**   |  |  |  | | --- | --- | --- | | **Assessment** | **Weight** | **Date** | | Homework assignments | 30% | Throughout semester | | Midterm Project | 20% | TBA | | Final Project | 50% | TBA | | Total | 100% |  | |  |  |  | |
| **M** | **Explanation of Assessments** | There will be regular homework assignments, midterm exam and final exam / project. |
| **N** | **Student Academic Integrity Code Statement** | All students are expected to abide by the Student Academic Integrity Code as articulated in the AUS graduate catalog. |

**Contribution of Course to Program Outcomes:** This course contributes in a significant way to the accomplishment of the following program outcomes:

* Apply advanced mathematical analysis to mathematical models
* Employ mathematical methods to model and solve practical problems
* Conduct independent research in specialized areas of mathematics
* Formulate problems in mathematical terms arising in related areas such as engineering, finance and the natural and physical sciences

**SCHEDULE**

|  |  |  |
| --- | --- | --- |
| **WEEK Number** | **CHAPTER** | **NOTES** |
| 1 | * Introduction * Basic definitions * Linear regression * Gradient descent * Normal equations | * Introduction * Lecture 1 |
| 2 | * Locally weighted regression * Linear regression: probabilistic interpretation * Logistic regression * Perceptron | * Lecture 2 |
| 3 | * Newton's method * Exponential family * Generalized Linear Models (GLM) * Generative learning algorithms | * Lecture 3 |
| 4 | * Gaussians * Gaussian discriminant analysis * Generative vs Discriminant comparison * Naive Bayes * Laplace Smoothing | * Lecture 4 |
| 5 | * Event models * Neural networks | * Lecture 5 |
| 6 | * Support vector machines: intuition * Primal / dual optimization problem (KKT) * SVM dual * Kernels * Soft margin * SMO algorithm | * Lecture 6 |
| 7 | * Generalized additive models (GAM) * Tree-based methods * Boosting * Boosting trees | * Lecture 7 |
| 8 | Case studies 1 | * **Midterm Exam** |
| 9 | * Learning theory * Bias / variance * Empirical Risk Minimization (ERM) * Union Bound / Hoeffdiny inequality * Uniform convergence * VC dimension | * Lecture 8 |
| 10 | * Model selection * Cross validation * Feature selection * Bayesian statistics vs regularization * Online learning | * Lecture 9 |
| 11 | * Unsupervised learning * Clustering (k-means) * Mixture of Gaussians * Jensen's inequality * EM (expectation – maximization) * Mixture of non Bayes * Factor analysis | * Lecture 10 |
| 12 | * Factor analysis: EM steps * Principal Component analysis (PCA) * Latent Semantic Indexing (LSI) * Singular Value Decomposition (SVD) * Independent Component Analysis (ICA) | * Lecture 11 |
| 13 | * Recommender systems * Factorization Machines | * Lecture 12 |
| 14 | Case studies 2 |  |
| 15 | Project presentations | * **Final Exam** |